基于突变指数的论文创新性识别方法研究*

陆凯余　　杨雪

中山大学信息管理学院 广州 510006

摘 要 [目的/意义] 构建突变指数测度论文创新性，为完善学术论文创新性评价体系提供参考。**[方法/过程]** 使用大型语言模型（LLM）Mistral-openorca 抽取论文研究问题和研究方法，以论文研究问题和研究方法与已有研究的相似性及其引起后续研究的追随程度两维度构建突变指数对论文创新性进行测度。**[结果/结论]** 高创新性得分的论文通常具有较高的被引频次；创新性排名位于前 10 名的论文中有 7 篇论文提出了原创性的方法或工具，而创新性排名位于后 10 名的论文研究主题和方法在论文发表时都较为成熟，创新程度较低；论文创新性得分与论文被引频次之间的相关系数为 0.530，实证研究结果进一步证实了本文提出的突变指数在测度论文创新性方面的有效性。

关键词：科技创新 LLM 突变指数 论文创新性评价

分类号：G301

创新是推动科学进步的重要驱动力，近年来，我国大力推进创新驱动发展战略，出台系列政策鼓励科技创新发展。习近平总书记在党的二十大报告[1]中指出必须坚持创新是第一动力，深入实施创新驱动发展战略，突出强调了科技创新对于建设科技强国的重要作用。《中共中央关于制定国民经济和社会发展第十四个五年规划和二〇三五年远景目标的建议》[2]中，将“坚持创新”列为未来五年十二项重要领域工作的首位，指出在科技创新上要强化国家战略科技力量，激发人才创新活力，完善科技创新体制机制，坚决破除“唯学历、唯职称、唯学历、唯奖项”，在此背景下，科技成果创新性评价成为科技创新发展过程中面临的关键问题。

科技论文是科技成果的重要形式，其承载了各领域的前沿研究成果，是国家整体科技创新水平的体现，在新时代的背景下如何构建一套科学的科技论文创新评价体系，推动科技论文创新性识别研究迈向新台阶是当前学界与实践界共同关注的重要问题，而设计科学、行之有效的科技论文创新性识别方法则是开展科技论文创新性评价活动的基础与核心，同时也是解决当前科技论文创新性评价与识别领

[作者简介] 陆凯余，中山大学信息管理学院本科生，通信作者 E-mail：andre_lukaiyu@foxmail.com；杨雪，中山大学信息管理学院硕士研究生。

域问题的关键。然而当前学界在学术论文创新性识别方法的研究方面依旧留有探索空间，因此，本文在已有研究基础上，构建突变指数测度论文创新性，以期为完善科技论文创新性评价方法体系提供一定参考。

1 相关概念与研究现状/ Relevant Concepts and Research Status

1.1 论文创新性的概念

学界对创新性的定义和理解尚未形成统一标准，但普遍认同的是，创新性论文应当在某种程度上推动科学知识的发展和进步。根据 M. Saunila 等[3]的研究，创新性可以被视为科研成果在理论、方法或实践层面上对现有知识体系的产生贡献，不仅体现在研究成果的新颖性上，还体现在其对学术界或实践领域的深远影响上。因此，创新性应当包括两个基本要素：新颖性（Novelty）和有用性（Usefulness）[4]。新颖性指的是研究成果在学术界前所未有，而有用性则指这些成果能够解决实际问题或推动理论发展。H.F. Moed [5]认为，创新性可以通过论文的引用情况间接衡量，即一篇论文如果能在短时间内被广泛使用，那么它很可能具有较高的创新性。所以本文认为，论文创新性可以通过论文是否有新的知识提出和是否为后续研究带来价值两个维度进行量化。

1.2 论文创新性的定量测度方法

当前研究通常从论文新颖性和有用性两个维度对论文创新性进行测度[6]。新颖性的知识元素是衡量论文是否创新的首要前提条件，也是检验论文是否具备潜在创新价值的标准之一，具体测度时，学界通常通过差异度对科技论文的新颖性进行表征，从论文中是否出现新的知识元素及知识元素组合的视角进行测度[7]。但受限于实体识别和提取的难度，早期许多研究多是基于论文主题词、关键词以及参考文献等指标及其组合评价论文创新性，上述指标实际在刻画论文核心内容方面的作用较为有限，而当前深入到论文核心知识内容层面的相关测度指标的开发依旧有待探索。通过对学界相关研究的梳理可见，研究问题和方法可视为科技论文的核心组成部分，已有研究表明新颖的问题或方法意味科技论文具备创新潜力，更有可能带来创新[8]。因此，从论文研究问题与研究方法出发识别创新性论文是较为有效的方式，本文也将采取基于论文研究问题与方法组合相似度的方式对论文新颖性进行量化。

有用性是衡量学术论文价值的关键指标之一，目前，研究人员普遍采用被引频

次作为衡量学术论文有用性的主要工具。然而，被引频次作为衡量有用性的工具存在一定的局限性，引用次数可能受到人为操纵；其次，被引频次可能无法全面反映论文的实用价值，因为一些具有长期影响的研究可能在初期并未获得大量的引用；此外，被引频次也忽视了引用的质量，即引用的正面或负面评价，以及论文对特定领域或问题的深入探讨程度。本文在现有研究基础上，通过当前研究的研究主题与方法引起后续研究追随程度的方式测度论文有用性，以此构建突变指数测度论文创新性。

1.3 实体抽取研究

在实体识别和抽取方面，现有研究多使用 SVM 等传统机器学习模型或深度学习模型、BERT 等语言模型，但上述模型都对数据标注有较高的依赖，在标注数据较少或者质量不佳时模型的性能会受到较大的影响，同时部分研究在抽取问题和方法实体时还面临着人工标注成本过高、基于简单规则匹配抽取实体带来的准确率问题等挑战，一个平衡成本和准确率的问题、方法实体抽取方法成为当前亟需解决的问题。近年来，ChatGPT 等大语言模型快速发展，其已在阅读理解和情感分析等自然语言处理任务中表现出较好的性能，同时已有研究将大语言模型用于实体识别任务，如 M.P. Polak 等[9]通过 ChatExtract 方法在不需要额外调优和进行大量代码开发工作的情况下即可识别相应实体，ChatGPT4 等较好的 LLMs 在应用于实体提取的测试中达到了接近 90%的精确率和召回率。总之，大语言模型在实体抽取的效果上显著优于其他模型，因此，本文用大语言模型进行实体抽取。

当前研究虽然对论文创新性测度方法展开了系列探索，但仍留有较多探索空间，具体表现为测度指标未完全反映论文核心研究内容与问题、实体抽取效率及准确率低、衡量创新性的方法不完善等。基于上述学者对创新性的定义以及本文对创新性概念的理解，本文进一步构建包含新颖性与有用性两维度的突变指数的论文创新性识别模型测度论文创新性。

2 论文创新性识别方法设计/ Design of Innovative Identification Methods for Academic Papers

依据已有研究，论文P_i的创新性可通过该论文自身新颖性及有用性衡量，本文使用能够体现论文核心内容的研究主题与方法的来衡量论文是否具备新颖性和有用性。新颖性具体可通过抽取的主题词或方法词在P_i发表前的词频来测度，有用性

则使用文章P_i发表前后对应主题词与方法词的词频增长速度变化来测度，具体设计如下规则：

（1）新颖性：如果P_i发表前该词的词频较低，则认为该词是新颖的，但如果相关词在P_i发表前拥有较高的词频，但是在P_i发表前的数年内词频较低，则也认为该词是新颖的。

（2）有用性：如果文章P_i发表后，其主题词或方法词的增长速度明显增大，则认为该词能够引发后续研究的追随[10]，则其具有较高的有用性。同时，如果文章P_i提出的方法被后来文章所使用，那么该方法具有更高的有用性（相对于第一种情况）。

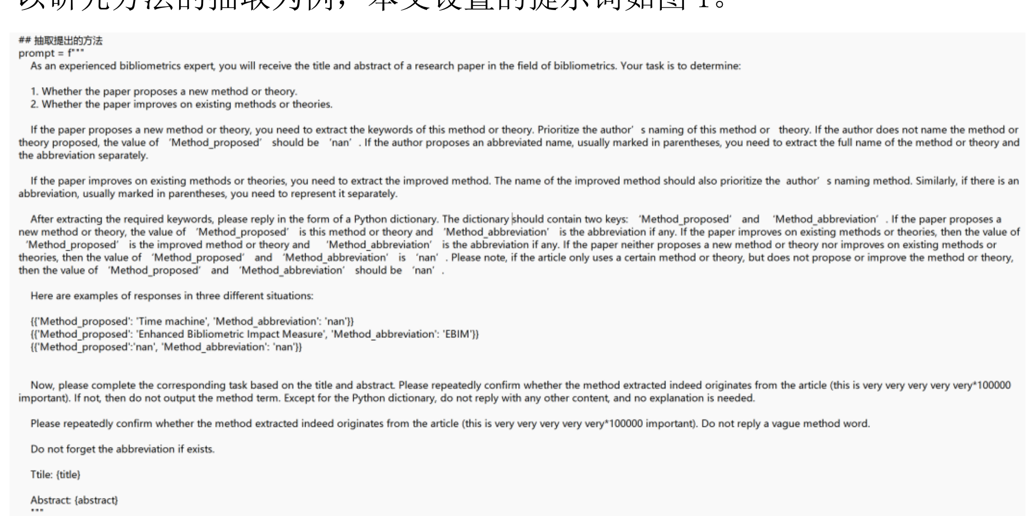
本文首先基于 LLM 模型进行问题词和方法词的抽取，对抽取的实体进行实体对齐的基础上依据本文设计的突变指数测度论文创新性。

2.1 论文主题词和方法词抽取

本文使用大语言模型抽取论文研究主题词与研究方法词，具体调用的模型为 Mistral-7B-OpenOrca，该版本模型是一个由 Mistral-7B 基础模型在 OpenOrca 数据集和 Mistral 7B 的基础上进行微调的语言模型。在发布时在所有小于 30B 的模型中排名第一，其使用 OpenChat 打包，与 Axolotl 一起训练。能够达到 Llama2-70B-chat 性能的 98%。是第一个具有突破性性能的完全开放模型，即使在适度的消费者 GPU 上也能完全加速运行。因此，本文选取该模型进行实体抽取。

本文编写 python 自定程序，将提示词和每篇文献的标题与摘要传入大语言模型 Mistral-Openorca，以获得结构化的主题词和方法词数据。

以研究方法的抽取为例，本文设置的提示词如图 1。



```
## 抽取提出的方法
prompt = f"""
   As an experienced bibliometrics expert, you will receive the title and abstract of a research paper in the field of bibliometrics. Your task is to determine:

   1. Whether the paper proposes a new method or theory.
   2. Whether the paper improves on existing methods or theories.

   If the paper proposes a new method or theory, you need to extract the keywords of this method or theory. Prioritize the author’s naming of this method or theory. If the author does not name the method or theory proposed, the value of ‘Method_proposed’ should be ‘nan’. If the author proposes an abbreviated name, usually marked in parentheses, you need to extract the full name of the method or theory and the abbreviation separately.

   If the paper improves on existing methods or theories, you need to extract the improved method. The name of the improved method should also prioritize the author’s naming method. Similarly, if there is an abbreviation, usually marked in parentheses, you need to represent it separately.

   After extracting the required keywords, please reply in the form of a Python dictionary. The dictionary should contain two keys: ‘Method_proposed’ and ‘Method_abbreviation’. If the paper proposes a new method or theory, the value of ‘Method_proposed’ is this method or theory and ‘Method_abbreviation’ is the abbreviation if any. If the paper improves on existing methods or theories, then the value of ‘Method_proposed’ is the improved method or theory and ‘Method_abbreviation’ is the abbreviation if any. If the paper neither proposes a new method or theory nor improves on existing methods or theories, then the value of ‘Method_proposed’ and ‘Method_abbreviation’ is ‘nan’. Please note, if the article only uses a certain method or theory, but does not propose or improve the method or theory, then the value of ‘Method_proposed’ and ‘Method_abbreviation’ should be ‘nan’.

   Here are examples of responses in three different situations:

   {{'Method_proposed': 'Time machine', 'Method_abbreviation': 'nan'}}
   {{'Method_proposed': 'Enhanced Bibliometric Impact Measure', 'Method_abbreviation': 'EBIM'}}
   {{'Method_proposed':'nan', 'Method_abbreviation': 'nan'}}


   Now, please complete the corresponding task based on the title and abstract. Please repeatedly confirm whether the method extracted indeed originates from the article (this is very very very very very*100000 important). If not, then do not output the method term. Except for the Python dictionary, do not reply with any other content, and no explanation is needed.

   Please repeatedly confirm whether the method extracted indeed originates from the article (this is very very very very very*100000 important). Do not reply a vague method word.

   Do not forget the abbreviation if exists.

   Ttile: {title}

   Abstract: {abstract}
"""
```

图 1 抽取研究方法的提示词

Figure 1 Prompt for extracting research methods

方法词抽取过程如下：

（1）让 LLM 识别出提出新方法的论文，抽取出相应的方法词作为第一批方法词集合M_1，并获取M_1的基于语义的词嵌入向量。

（2）让 LLM 抽取所有论文所用研究方法的方法词（无论该文章是否提出了新方法），作为第二批方法词集合M_2，并获取M_2基于语义的词嵌入向量。

（3）让 LLM 再次对所有文章进行方法词抽取，但是，此次对于第i篇文章P_i进行方法词抽取时，需要先根据论文P_i在M_2中的方法词$M_{2,i}$通过 KNN 算法利用提前获取到的词嵌入查找出$M_{2,i}$在M_1中最相似的10个词汇作为备选词发送给 LLM，并提示 LLM 如果备选词中存在P_i文章所使用的方法，则直接输出相应的方法词，反之从摘要中抽取新的方法词，并将新抽取到的方法词放入集合M_1并获取新方法词的词嵌入向量，获得第三批方法词集合M_3。

主题词抽取过程如下：

（1）让 LLM 抽取所有文章的研究主题词，作为第一批主题词集合T_1，并获取T_1基于语义的词嵌入向量。

（2）筛选出主题词集合T_1中所有词频大于等于5的主题词作为第二批主题词集合T_2。

（3）与上述方法词的抽取过程同理，让 LLM 再次对所有文章进行主题词抽取，但是，此次对于第i篇文章P_i进行主题词抽取时，先根据P_i文章在T_1中的主题词$T_{1,i}$通过 KNN 算法利用提前获取到的词嵌入查找出$T_{1,i}$在T_2中最相似的10个词汇作为备选词发送给 LLM，并提示 LLM 如果备选词中存在P_i文章的研究主题，则直接输出相应的主题词，反之从标题和摘要中抽取新的主题词，并将新抽取到的主题词放入集合T_2并获取新主题词的词嵌入向量。获得第三批主题词集合T_3。通过上述抽取过程即可获得研究所用主题词与方法词集。

2.2 实体对齐

在抽取研究主题词与研究方法词的基础上，需对主题词与方法词进行实体对齐以保证计算结果的准确性。具体而言，本文利用 Mistral-Openorca 词嵌入模型，事先获取第一批高频主题词和方法词的词嵌入，随后利用 KNN 算法查找出与每篇文献的主题和方法最相似的 10 个备选词传入大语言模型，得到对齐后的主题词和

方法词。

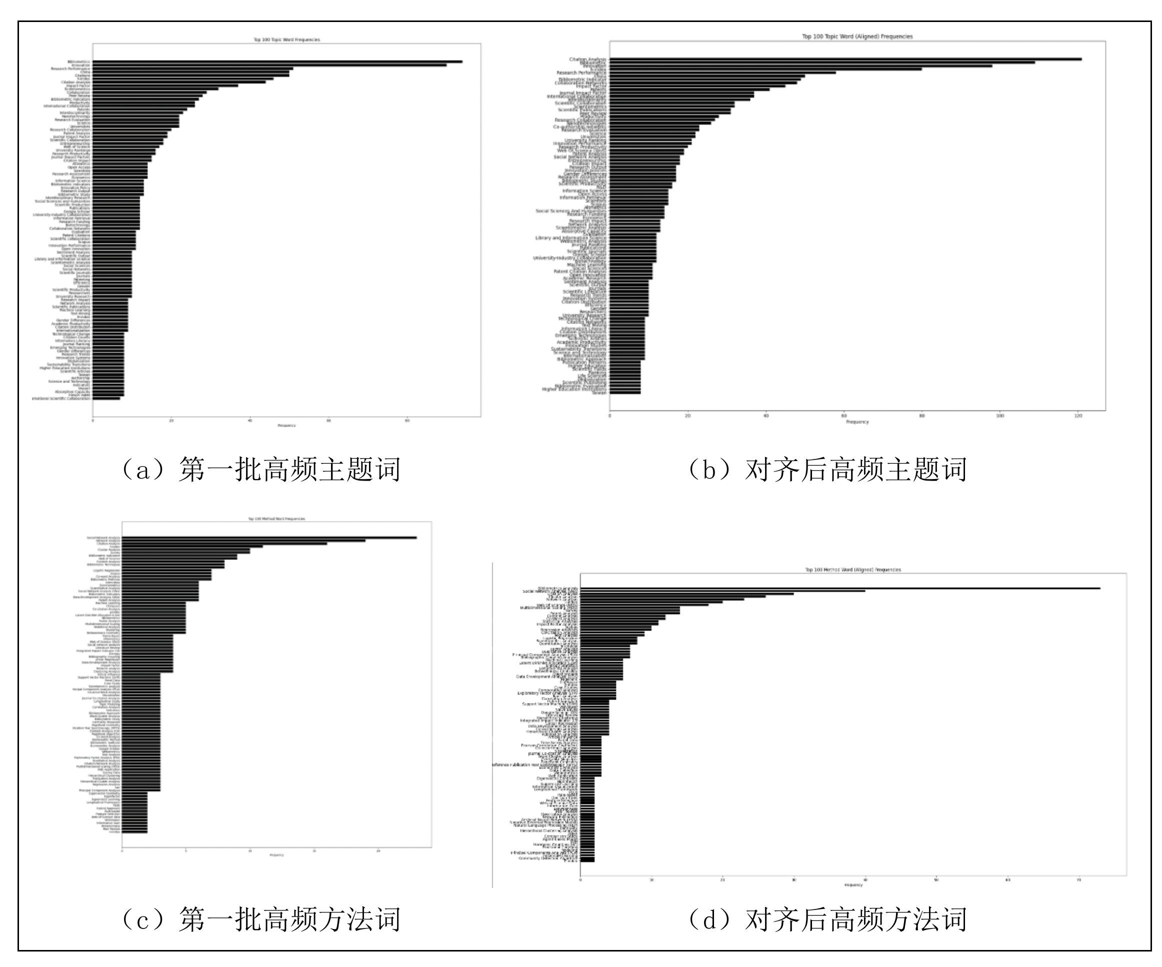
最终获得第一批主题词共有 18143 个，第一批方法词共 8674 个；对齐后主题词共 15121 个，方法词共 6121 个，具体主题抽取与对齐结果示例如图 2 所示。除此，本文也对实体对齐前后的主题词和方法词的词频分布图进行了分析，见图 3。通过对齐前后的词频分布可以看出，本文所用实体对齐方式能够将语义基本一致的主题词和方法词进行统一，如第一批高频主题词中 Citations 和 Citation Analysis 都对应引文分析的主题，在对齐后统一为 Citation Analysis，Citations 则未出现在对齐后的高频词中；对齐前的高频方法词中，Bibliometric Method 和 Bibliometric Approach 实际都是计量学方法，其对齐后都统一为 Bibliometric Analysis。



Article Title	Abstract	methods_proposed_new	topics	topics_aligned	methods	methods_aligned
Software survey	We preser	{'Method_proposed': 'VOSviewer', 'Method_abbre	['Bibliometric Mapping', 'Softwar	['Bibliometric Mapping', 'Software']	[]	[]
A new approach	A size-ir	{'Method_proposed': 'SCImago Journal Rank indi	['SCImago Journal Rank (SJR)', 'S	['SCImago Journal Rank (SJR)', 'Scholarly	['SCImago Journal Rank (SJR) Indicato	['SCImago Journal Rank (SJR) Indica
Measuring conte	This pape	{'Method_proposed': 'Source Normalized Impact	['Journal Citation Impact', 'SNIF	['Journal Citation Impact', 'SNIP']	[]	[]
Is science becc	In the la	{'Method_proposed': 'Integration score', 'Meth	['Interdisciplinarity', 'Research	['Interdisciplinarity', 'Research Fields']	['Bibliometric Indicators', 'Integrat	['Bibliometrics Analysis', 'Integra
Introducing Cit	We introc	{'Method_proposed': 'CitedReferencesExplorer (	['Bibliometrics', 'Reference Publ	['Bibliometric', 'Reference Publication Ye	[]	[]
Bibliometric in	In this p	{'Method_proposed': 'CITAN', 'Method_abbreviat	['Bibliometric impact assessment'	['Bibliometric Impact Assessment', 'R']	['CITAN package']	['CITAN Package']
hg-index: a new	To be ab	{'Method_proposed': 'hg-index', 'Method_abbrev	['h-index', 'Scientific Output']	['h-index', 'Scientific Output']	[]	[]
Evaluating altn	The rise of the social web and its uptake by scholars h		['Altmetrics', 'Evaluation']	['Altmetrics', 'Evaluation']	['Correlation Tests', 'Content Analys	['Correlation Tests', 'Content Ana
Sentiment analy	Sentiment analysis is an important current research are		['Sentiment Analysis', 'Combined	['Sentiment Analysis']	['Rule-based Classification', 'Superv	['Rule-based Classification', 'Sup
Using altmetric	The prospects of altmetrics are especially encouraging		['Altmetrics', 'Research Impact',	['Altmetrics', 'Research Impact', 'Humanit	['Mendeley', 'Twitter', 'Library Thin	['Mendeley', 'Twitter', 'Library Th
Methodi Ordinat	An increa	{'Method_proposed': 'Methodi Ordinatio', 'Meth	['Methodi Ordinatio', 'Scientific	['Methodi Ordinatio', 'Scientific Publicat	['ProKnow-C (PKC)', 'InOrdinatio (IOR	['ProKnow-C (PKC)', 'InOrdinatio (
Which percentil	For compa	{'Method_proposed': 'P100', 'Method_abbreviati	['Percentile-based approach', 'Nc	['Percentile-based Approach', 'Normalized	['P100', 'SCImago group', 'Centre for	['P100', 'SCImago Group', 'Centre
ASIRPA: A compr	ASIRPA is	{'Method_proposed': 'ASIRPA', 'Method_abbrevia	['Comprehensive Approach', 'Socie	['Comprehensive Approach', 'Societal Impac	[]	[]
Interactive ove	Recent ac	{'Method_proposed': 'Interactive overlays', 'M	['Interactive overlays', 'Global	['Interactive Overlays', 'Global Journal M	['Web-of-Science data', 'Spring-embed	['Web-of-Science Data', 'Spring-em
Mapping the res	Research on aquaculture is expanding along with the exc		['Aquaculture', 'Bibliometrics']	['Aquaculture', 'Bibliometric']	['Latent Semantic Analysis', 'Topic M	['Latent Semantic Analysis', 'Topi
The Z-index: A	We preser	{'Method_proposed': 'Z-index', 'Method_abbrevi	['Z-index', 'Productivity and Imp	['Z-index', 'Productivity And Impact']	[]	[]
CitNetExplorer:	We preser	{'Method_proposed': 'CitNetExplorer', 'Method_	['CitNetExplorer', 'Citation Netw	['CitNetExplorer', 'Citation Networks']	[]	[]
Referenced Publ	We have c	{'Method_proposed': 'Referenced Publication Ye	['Referenced Publication Years Sp	['Referenced Publication Years Spectroscop	['RPYS (Routine for Referenced Publica	['RPYS (Routine For Referenced Pub
Can alternative	This study compares Spanish and UK research in eight su		['Bibliometrics', 'Citation Count	['Bibliometric', 'Citation Analysis']	['Webometric Analyst', 'Altmetric.com	['Webometric Analyst', 'Altmetric.
Growth and stru	We study the evolution of Slovenia's scientific collabo		['Slovenia', 'Scientific Collabor	['Slovenia', 'Scientific Collaboration Net	['Publication Records', 'Yearly Resol	['Publication Records', 'Yearly Re
Aspect-based se	In this s	{'Method_proposed': 'Aspect-based sentiment an	['Movie Reviews', 'Sentiment Anal	['Movie Reviews', 'Sentiment Analysis']	[]	[]
Citation impact	Research	{'Method_proposed': 'Stepwise Regression Analy	['Citation Impact Prediction', 'S	['Citation Impact Prediction', 'Scientific	['Stepwise Regression Analysis']	['Stepwise Regression Analysis']
Detecting hot t	Twitter i	{'Method_proposed': 'MVTD', 'Method_abbreviati	['Multiview', 'Topic Detection',	['Multiview', 'Topic Detection', 'Twitter'	['MVTD (Multiview Twitter Detection)',	['MVTD (Multiview Twitter Detectio
Using neural ne	This study applies the artificial neural network techni		['Patent Performance', 'Market Va	['Patent Performance', 'Market Value', 'Ph	['Artificial Neural Network (ANN)', 'I	['Artificial Neural Network (ANN)',
The impact of m	Different from previous studies analyzing the influence		['Multilevel Networks', 'Innovati	['Multilevel Networks', 'Innovation']	['Panel Data', 'Negative Binomial Reg	['Panel Data', 'Negative Binomial
q2-Index: Quant	Bibliome	{'Method_proposed': 'q2-Index', 'Method_abbrev	['q2-Index', 'Bibliometric studie	['Q2-Index', 'Bibliometric']	['Hirsch core', 'h-index', 'g-index',	['Hirsch Core', 'H-Index', 'G-inde
Clustering scie	Topic modeling is a type of statistical model for disco		['Topic Modeling', 'Clustering Sc	['Topic Modeling', 'Clustering Scientific	['Latent Dirichlet Allocation (LDA)',	['Latent Dirichlet Allocation (LDA
LDA-based onlin	In the in	{'Method_proposed': 'Online Topic Detection us	['Online Topic Detection', 'Tensc	['Online Topic Detection', 'Tensor Factori	['OTD-TF (Online Topic Detection usin	['OTD-TF (Online Topic Detection U
An author co-ci	Author co-citation analysis (ACA) is an important metho		['Author Co-citation Analysis', '	['Author Co-citation Analysis', 'Informati	['ACA', 'Factor Analysis', 'Pajek']	['ACA', 'Factor Analysis', 'Pajek'
Quantifying the	While the modern science is characterized by an exponen		['Quantifying', 'Cognitive Extent	['Quantifying', 'Cognitive Extent']	['Big Data Method', 'Lexical Diversit	['Big Data Method', 'Lexical Diver
Normalization c	A differe	{'Method_proposed': 'Mean Normalized Reader Sc	['Normalization of Mendeley reade	['Normalization Of Mendeley Reader Counts'	[]	[]
SRank: Shortest	Similari	{'Method_proposed': 'SRank', 'Method_abbreviat	['Similarity Estimation', 'Shorte	['Similarity Estimation', 'Shortest Paths'	['SRank (S-Rank)']	['SRank (S-Rank)']
A study of the	Sentiment analysis has drawn considerable interest amon		['Sentiment Analysis', 'Arabic Te	['Sentiment Analysis', 'Arabic Text']	['Stemming', 'Feature Correlation', 'I	['Stemming', 'Feature Correlation',
Time gap analys	This study proposes a temporal analysis method to utili		['Time Gap Analysis', 'Topic Mode	['Time Gap Analysis', 'Topic Model']	['Temporal Technique', 'Latent Dirich	['Temporal Technique', 'Latent Dir
Harmonic public	Bibliome	{'Method_proposed': 'Harmonic Counting', 'Meth	['Bibliometric counting methods',	['Bibliometric Counting Methods', 'Authors	['Harmonic counting (HC)', 'Arithmeti	['Harmonic Counting (HC)', 'Arithm
Some new tests	A centra	{'Method_proposed': 'tf*idf formula', 'Method_	['Relevance Theory', 'Informatior	['Relevance Theory', 'Information Science'	['Term Frequency-Inverse Document Fre	['Term Frequency-Inverse Document
A research impa	This pape	{'Method_proposed': '(n)h(3)', 'Method_abbrevi	['Impact Indicator', 'Institutior	['Impact Indicator', 'Institutional Resear	['nh(3)', 'Mean Citation Rate']	['Nh(3)', 'Mean Citation Rate']
Crowdsourcing,	We show through an analysis of a massive data set from		['Crowdsourcing', 'Attention', 'F	['Crowdsourcing', 'Attention', 'Productivi	[]	[]
Analysis of res	E-commerce (EC) is sweep across the globe and has becom		['E-commerce', 'Text Mining']	['E-commerce', 'Text Mining']	['Multidimensional Scaling', 'Knowled	['Multidimensional Scaling (MDS)',
Spatial scientc	We propo	{'Method_proposed': 'Spatial scientometrics',	['Spatial Scientometrics', 'Cumul	['Spatial Scientometrics', 'Cumulative Res	[]	[]
Network collabc	This paper aims to explore the role of each country in		['Network collaboration', '6th Fr	['Network Collaboration', '6th Framework P	['Social network analysis (SNA)', 'Tw	['Social Network Analysis (SNA)', '
Empirical study	A knowle	{'Method_proposed': 'Latent Dirichlet Allocati	['Patent Documents', 'Knowledge C	['Patent Documents', 'Knowledge Organizati	['Term Clumping', 'Latent Dirichlet A	['Term Clumping', 'Latent Dirichle
Visualization a	CiteSpace is a visual document analysis software, by wh		['Visualization', 'Quantitative s	['Visualization', 'Quantitative Study', 'B	['CiteSpace (CS)', 'Web of Science (W	['CiteSpace (CS)', 'Web Of Science
Identifying res	In this s	{'Method_proposed': 'Weighted Direct Citation	['Weighted direct citation links'	['Weighted Direct Citation Links', 'Librar	['Direct citations', 'Shared referenc	['Direct Citations', 'Shared Refer
Do altmetrics p	Today, it is not clear how the impact of research on ot		['Altmetrics', 'Societal Impact c	['Altmetrics', 'Societal Impact Of Researc	[]	[]
The representat	This article studies publication patterns in the social		['Publication Patterns', 'Social	['Publication Patterns', 'Social Sciences	['Comparative Analysis', 'Performance-	['Comparative Analysis', 'Performa
Efficient super	The disar	{'Method_proposed': 'Naive Bayes model', 'Meth	['Affiliations Disambiguation', '	['Affiliations Disambiguation', 'Bibliogra	['Naive Bayes Model', 'Semi-supervise	['Naive Bayes Model', 'Semi-superv
Applying the se	The Lattes Platform is a web-based system that brings t		['Semantic Web', 'Academic and Pr	['Semantic Web', 'Academic And Professiona	['Lattes Platform', 'Linked Data', 'F	['Lattes Platform', 'Linked Data',

图 2 主题词和方法词抽取与对齐结果

Figure 2 Extraction and alignment of topic and method keywords



（a）第一批高频主题词　　（b）对齐后高频主题词

（c）第一批高频方法词　　（d）对齐后高频方法词

图 3 实体对齐前后的词频分布情况

Figure 3 Word frequency distribution before and after entity alignment

2.3 突变指数计算

依据上述实体抽取结果和新颖性、有用性基本算法，本文构建了突变指数测度论文创新性。假如文章P_i在t_i日期发表，其共有$|T_i|$个主题词和$|M_i|$个方法词，其中主题词为：$T_{i,j(j=1,2,\ldots,|T_i|)}$，方法词为$M_{i,j(j=1,2,\ldots,|M_i|)}$。

在t_i日期前后window天数时间$[t_i - \mathrm{window}, t_i + \mathrm{window}]$中，主题词$T_{i,j}$在$P_i$发表前和发表后window天内的词频分别为$N_{post(T_{i,j})}$和$N_{pre(T_{i,j})}$；类似地，方法词$M_{i,j}$在$P_i$发表前和发表后window天内的词频分别为$N_{post(M_{i,j})}$和$M_{pre(T_{i,j})}$。$t_r$为同样使用了$T_{i,j}$或$M_{i,j}$的文章$P_r$的发表时间，$|t_i - t_r|$为$P_i$与$P_r$的发表日期差。$f(\Delta t)$和$g(\Delta t)$分别为主题词和方法词对突变指数的贡献的衰减函数，如果文章P_i与P_r拥有某一相同的主题词或方法词，那么P_i与P_r发表时间越接近，则该词对突变指数的贡

献越大。其中参数λ_1和λ_2为时间衰减参数，取值范围为(0,1)；ϕ为原创方法参数，如果$M_{i,j}$是原创方法，则ϕ取值$m(m > 1)$, 否则ϕ取值1。

所有参数window、λ_1、λ_2、m的具体数值根据实证领域的情况选取。

（1）主题$T_{i,j}$的突变指数：

$$\mathrm{Mutation}(T_{i,j}) = \frac{\sum_{r=1}^{N_{post}(T_{i,j})} f(|t_i - t_r|)}{\sum_{r=1}^{N_{pre}(T_{i,j})} f(|t_i - t_r|)} - \frac{\sum_{r=1}^{N_{pre}(T_{i,j})} f(|t_i - t_r|)}{\sum_{r=1}^{N_{post}(T_{i,j})} f(|t_i - t_r|)} \tag{1}$$

（2）方法$M_{I,j}$的突变指数：

$$\mathrm{Mutation}(M_{i,j}) = \frac{\sum_{r=1}^{N_{post}(M_{i,j})} g(|t_i - t_r|)}{\sum_{r=1}^{N_{pre}(M_{i,j})} g(|t_i - t_r|)} - \frac{\sum_{r=1}^{N_{pre}(M_{i,j})} g(|t_i - t_r|)}{\sum_{r=1}^{N_{post}(M_{i,j})} g(|t_i - t_r|)} \tag{2}$$

（3）文章P_i的突变指数：

$$\mathrm{Mutation}(P_i) = \sum_{j=1}^{|T_i|} Mutation(T_{i,j}) + \sum_{j=1}^{|M_i|} Mutation(M_{i,j}) \tag{3}$$

（4）衰减函数$\mathrm{f}(\Delta t)$和$\mathrm{g}(\Delta t)$：

$$\mathrm{f}(\Delta t) = \exp(-\lambda_1 * \Delta \mathrm{t}) \tag{4}$$

$$\mathrm{g}(\Delta t) = \phi * \exp(-\lambda_2 * \Delta t) \tag{5}$$

3 实证研究/ Empirical Research

3.1 数据收集与处理

本文所用数据来源于 Web of Science（WoS 数据库)，为便于结果解读与分析，本文选取本专业中信息计量学方向研究论文作为样本数据。具体而言，在 WoS 核心合集数据库中，通过期刊名称（publication titles）进行检索，选取 SCIENTOMETRICS、JOURNAL OF INFORMETRICS、Research Evaluation、Information Processing and Management、Journal of Information Science、Research Policy 等 6 本与计量学相关的期刊。由于上述期刊中早期发表的论文较少，而近几年的文章的后向文章数量较少，为保证有足够的观察时间窗口，同时确保数据的有效性，减少出现异常值的频率，本文选取 2009 年至 2016 年之间的文献进行实证分析。本文获取上述期刊在 2009-2016 年发表的所有论文的全记录（Full record）与引用数据。由于初次获取的样本数据存在标题、摘要等数据项不全的情况，因此

本文对该部分数据进行了清洗，清洗后共计剩余样本论文 14,138 篇，其中用于实证分析的 2009 年至 2016 年之间的样本论文共 4881 篇。

3.2 结果分析

3.2.1 总体得分分布情况

论文基本统计结果见表 1，通过表 1 可见，论文创新性均值为 1.043，中位数为 0.2002，最大值为 208.693，最小值为-12.948，得分大于中位数的论文有 2441 篇，小于中位数的论文有 2440 篇。论文创新性测度结果如图 4 所示，图中横坐标为创新性分数，纵坐标为分布在特定区间中的样本的概率，越高表示分布在该区间中的样本概率越大。由图 4 可见，论文创新性集中分布在区间（-3，2]和区间（2，7]中。

表 1 测度结果描述性分析

Table 1 Descriptive analysis of measurement results

均值	中位数	标准差	最大值	最小值	25%	50%	75%
1.043	0	4.775	208.693	-12.948	0.0	0.200	1.613

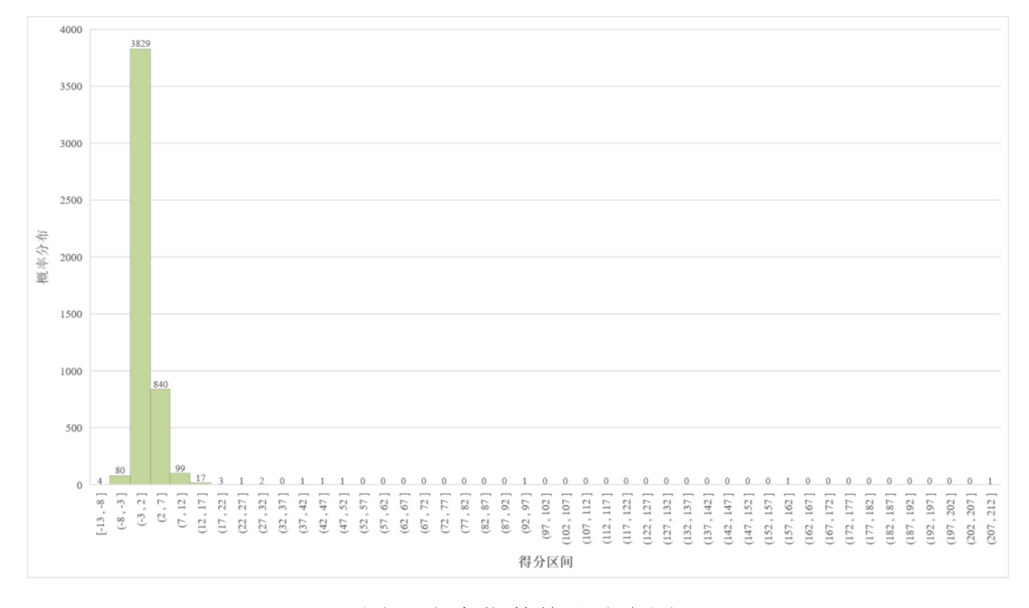


图 4 突变指数统计分布图

Figure 4 Statistical Distribution of Mutation Index

3.2.2 排名最高与最低的论文比较分析

本文进一步比较分析样本数据中创新性得分位于前 10 名与后 10 名的论文以检验测度模型与测度方法的有效性。分析结果见表 2 和表 3。表 2 为创新性最强的 10 篇论文。本文对上述论文进行了逐篇分析，对其是否提出新颖的研究方法与研

究主题等内容进行分析。通过表 6 中创新性最强的 10 篇论文的基本情况可见，突变指数前 10 篇文章中前 7 篇皆提出了原创方法，具有较高的创新性，如排名位于第 1 名的论文：Software survey: VOSviewer, a computer program for bibliometric mapping，该论文提出了 VOSviewer 计量工具，作者是 NJ. van Eck 等人；排名第 2 的论文：A new approach to the metric of journals' scientific prestige: The SJR indicator，该论文提出了一种衡量科学期刊影响力的指标 SCImago Journal Rank indicator，作者是 González-Pereira 等人，SJR 考虑了期刊文章被引用的质量和来源，而不仅仅是引用的数量，从而提供了一个更为全面和细致的期刊评价体系。排名第 3 的论文“Measuring contextual citation impact of scientific journals”由 Moed 等人撰写，提出了 Source Normalized Impact per Paper (SNIP)指标。SNIP 旨在衡量期刊的引用影响，考虑了学科领域的引用行为差异，为期刊评价提供了新的视角。排名第 4 的论文“Introducing CitedReferencesExplorer (CRExplorer): A program for reference publication year spectroscopy with cited references standardization”由 Thor 等人撰写，介绍了 CRExplorer 这一工具，它用于分析参考文献的出版年份，并通过标准化被引用参考文献的出版年份来增强对研究趋势的理解。排名第 5 的论文“Bibliometric impact assessment with R and the CITAN package”由 M.Gagolewski 撰写，提出了 CITAN 包，这是一个基于 R 语言的工具，用于进行科学计量学影响评估，为研究者提供了强大的分析工具。排名第 6 的论文“hg-index: a new index to characterize the scientific output of researchers based on the h- and g-indices”由 S.Alonso 等人撰写，提出了 hg-index，这是一个新的指标，结合了 h-index 和 g-index，旨在更全面地描述研究人员的科学产出。这些论文的共同特点是提出了新的工具、方法或指标，这些都对科学计量学领域的发展产生了重要影响。

同时通过分析发现，虽然论文 Introducing CitedReferencesExplorer (CRExplorer): A program for reference publication year spectroscopy with cited references standardization 和 Bibliometric impact assessment with R and the CITAN package 等的被引频次相对而言并不高，但通过分析可见，这两篇分别提出 CRExplorer 和 CITAN package 工具，它们都是计量学领域非常实用的工具，可见模型能够有效挖掘上述类型论文。

排名前 10 名的其他论文中，尽管某些论文可能没有提出全新的研究方法，但它们探讨的研究主题在其领域内具有新颖性和重要性。例如，P.Sud 等人的论文

“Evaluating altmetrics”发表于 2014 年，聚焦于 Altmetrics 这一新兴的科学计量学方法。在当时，Altmetrics 作为一种替代传统的引用计数的指标，能够提供关于学术成果在社交媒体和其他非传统渠道中的影响力和传播情况的见解。这项研究对于理解和评估学术成果的社会影响具有重要意义。B.Hammarfelt 的研究“Using altmetrics for assessing research impact in the humanities”同样关注 Altmetrics，特别是在人文学科领域的应用。这项研究探讨了如何利用 Altmetrics 来评估人文学科研究的影响力，这是一个相对较少被关注的领域，因此其研究主题具有创新性。另一方面，R.Prabowo 等人的论文“Sentiment analysis: A combined approach”发表于 2009 年，研究了情感分析的结合方法，包括基于规则的分类、监督学习和机器学习。情感分析是自然语言处理（NLP）领域的一个重要分支，它涉及到识别和提取文本中的情感倾向。在当时，机器学习和深度学习技术在情感分析中的应用还处于起步阶段，因此这项研究对于推动该领域的技术进步具有重要作用。这些论文虽然没有提出全新的方法，但它们在研究主题的选择上展现了创新性，为相关领域的研究提供了新的视角和思路。这些研究的高突变指数反映了它们在推动学术领域发展和创新方面的贡献。对上述论文的分析进一步验证了测度模型和方法在评估论文创新性方面的有效性。

得分位于最后 10 名的论文见表 3，可见排名后 10 名的论文缺乏研究方法和研究主题上的新颖性。例如，排名最低的论文“An informetric model for the success-index”由 F.Franceschini 等人在 2013 年发表，该论文没有提出新的方法，而是探讨了成功指数（success-index）的信息咨询模型（informetric model），尚未带来新的视角或突破，研究结果没有对现有的知识体系产生显著的补充或挑战。

表 2 突变指数排名前 10 的论文

Table 2 Top 10 Papers ranked by mutation index

作者	题目	被引量	提出方法	主题词	方法词	突变指数
NJ.van Eck 等[11],2010	Software survey: VOSviewer, a computer program for bibliometric mapping	7413	VOSviewer	Bibliometric Mapping, Software	\	208.694
González-Pereira 等[12]，2010	A new approach to the metric of journals' scientific prestige: The SJR indicator	402	SCImago Journal Rank indicator	SCImago Journal Rank (SJR), Scholarly Journals Prestige	SCImago Journal Rank (SJR) Indicator, Eigenvector Centrality	159.292
Moed 等[13]，2010	Measuring contextual citation impact of scientific journals	410	Source Normalized Impact per Paper	Journal Citation Impact, SNIP	\	95.235
AL.Porter 等[14]，2009	Is science becoming more interdisciplinary? Measuring and mapping six research fields over time	560	Integration score	Interdisciplinarity, Research Fields	Bibliometrics Analysis, Integration Score (Rao-Stirling Diversity), Science Mapping Visualization	49.319
A.Thor 等[15]，2016	Introducing CitedReferencesExplorer (CRExplorer): A program for reference publication year spectroscopy with cited references standardization	91	CitedReferencesExplorer (CRExplorer)	Bibliometric, Reference Publication Year Spectroscopy	\	43.662
M.Gagolewski 等[16]，2011	Bibliometric impact assessment with R and the CITAN package	45	CITAN	Bibliometric impact assessment, R	CITAN Package	41.433
S.Alonso 等[17]，2010	hg-index: a new index to characterize the scientific output of researchers based on the h- and g-indices	171	hg-index	h-index, Scientific Output	\	30.371
P.Sud 等[18]，2014	Evaluating altmetrics	202	\	Altmetrics, Evaluation	Correlation Tests, Content Analyses, Interview, Pragmatic Analyses	28.568
R.Prabowo,[19] 等，2009	Sentiment analysis: A combined approach	420	\	Sentiment Analysis	Rule-based Classification, Supervised Learning, Machine Learning	24.120
B.Hammarfelt 等[20]，2014	Using altmetrics for assessing research impact in the humanities	144	\	Altmetrics, Research Impact, Humanities	Mendeley, Twitter, Library Thing	20.704

表 3 突变指数排名后 10 的论文

Table 3 Papers ranked in the bottom 10 by mutation index

作者	题目	被引量	提出方法	主题词	方法词	突变指数
JM.van Zyl 等[21]，2013	A simulation study to investigate the accuracy of approximating averages of ratios using ratios of averages	0	\	Simulation Study, Accuracy, Averages Of Ratios, Ratios Of Averages	Zeta Distribution, Bootstrap Confidence Intervals, Log-Logistic Distribution	-5.770
D.Smith 等[22]，2013	What determines the location choice of R&D activities by multinational firms?	63	\	Multinational Firms, R&D Activities	\	-6.041
A.Pepe 等[23]，2010	Collaboration in sensor network research: an in-depth longitudinal analysis of assortative mixing patterns	36	\	Scientific Collaboration, Sensor Networks	\	-6.254
D.Bouyssou 等[24]，2014	An axiomatic approach to bibliometric rankings and indices	26	\	Bibliometric Rankings, Indices	\	-6.382
M.Song 等[25]，2015	Identifying the landscape of Alzheimer's disease research with network and content analysis	25	\	Alzheimers Disease, Research	Network Analysis, Content Analysis, Productivity Analysis	-6.798
J.Zheng 等[26]，2014	Influences of counting methods on country rankings: a perspective from patent analysis	9	\	Patent Analysis, Country Rankings	Whole Counting, Straight Counting, Normalized Counting, Complete-Normalized Counting	-6.842
M.Schreiber 等[27]，2014	Is the new citation-rank approach P100′ in bibliometrics really new?	3	\	Citation Rank, Bibliometric	P100]	-8.10
F.Franceschini 等[28]，2013	An informetric model for the success-index	3	\	Success-index, Informetric Model	Information Production Processes (IPPs), Theory Of Information Production Processes	-9.335
CY.Wu 等[29]，2014	Comparisons of technological innovation capabilities in the solar photovoltaic industries of Taiwan, China, and Korea	31	\	Technological Innovation Capabilities, Solar Photovoltaic Industries, Taiwan, China, Korea	Database Analysis, Patent Portfolio, Technology Platforms	-10.084
YW.Chang 等[30]，2014	Exploring scientific articles contributed by industries in Taiwan	8	\	Industries, Taiwan	\	-12.948

3.2.3 论文创新性与被引频次的相关性分析

上述分析表明本文构建的模型可有效识别出具有很强创新性的论文，本文进一步检验论文创新性与被引频次间的相关关系，如表 4。经检验，论文创新性分数与论文在核心合集中的被引频次呈显著正相关关系，相关系数为 0.530。

表 4 论文创新性与被引频次的相关关系

Table 4 The correlation between paper innovation and citation frequency

2009 年至 2016 年之间论文突变指数与被引频次的皮尔逊相关系数	P 值
0.530	<0.001

除此，本文也对不同创新性得分区间中论文被引频次的分布情况进行了分析，分析结果见图 5。图中的横坐标表示论文的创新性得分区间，而纵坐标则代表论文的平均被引频次。从图中可以观察到，创新性得分较高的论文区间通常伴随着较高的平均被引频次。具体来说，获得最高平均被引频次的论文集中在较高的创新性得分区间内。这表明，那些在创新性得分上位于顶端的论文，往往能够吸引更多的学术关注和引用。此外，图中的趋势还显示，随着论文创新性得分的提高，其被引频次也呈现上升趋势。这进一步印证了创新性与论文影响力之间的正相关关系。换句话说，那些在研究方法、理论框架或实践应用方面提出新观点和新思路的论文，更有可能成为学术界讨论和引用的焦点。

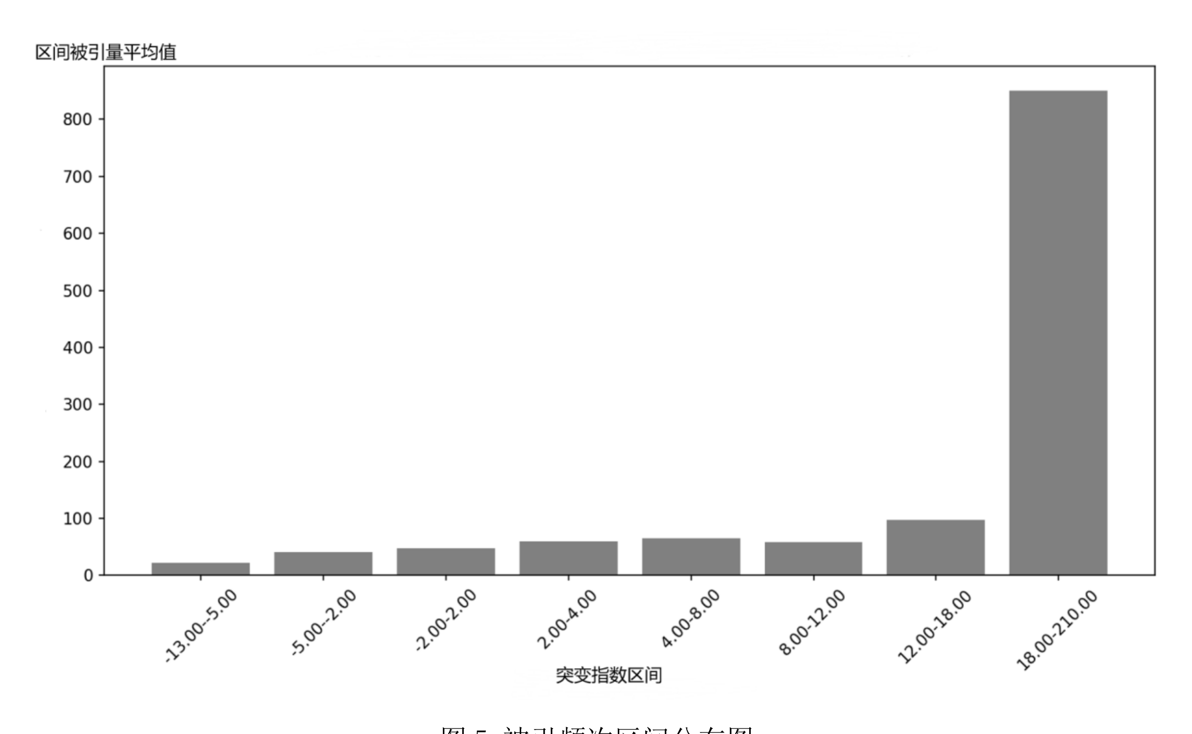


图 5 被引频次区间分布图

Figure 5 Distribution chart of citation frequency intervals

4 研究结果讨论与展望/ Discussion of findings and outlook

本文基于论文新颖性和有用性维度构建了测度论文创新性的突变指数，首先使用 LLM 抽取论文研究主题与研究方法，随后对抽取到的实体进行实体对齐，使用本文设计的突变指数测度论文创新性。实证研究结果表明在样本数据中，部分论文因其提出的新颖研究方法或主题而获得了较高的创新性得分，这些论文往往能够引领学术界的趋势，激发后续研究的灵感。然而，在研究方法或主题上未能展现出足够的新颖性或实用性的论文其创新性得分较低。此外，通过对排名靠前及靠后的论文的深入分析，本文发现高创新性得分的论文普遍提出了原创性较强的方法或理论，并且这些论文在学术界产生了较大的影响，这表明创新性评价指标与论文的学术影响力之间存在正相关关系。最后，本文还比较了不同创新性得分区间中论文的被引频次，结果显示创新性得分较高的论文平均被引频次也较高。

论文创新性评价对于学术界具有重要意义，它不仅能够指导研究者探索未知领域，还能够促进科研资源的合理分配和科研政策的有效制定。然而，当前的测度方法仍存在一些不足，例如对新兴研究主题的敏感度不足，以及对未命名原创方法的识别能力有限。实证研究表明，本文提出的基于 LLM 与突变指数的论文创新性识别方法，能够有效地识别出具有高创新性或突变性的论文，为科技论文的创新性评价提供了新的视角和工具。

本文也存在一定局限，主要体现为模型对于主题词或方法词的依赖性较强，对于未严格命名的原创方法或主题的识别能力有限。未来的研究可以在以下几个方向进行改进和拓展：首先，可以通过增加对未命名方法的使用情况的判断，提高模型对新兴研究主题的敏感度；其次，可以探索更细粒度的判断和对齐方法，以更好地识别和评价那些在学术界具有潜在影响力的创新性研究。此外，未来的研究还可以考虑结合其他学科的理论和方法，以进一步提高论文创新性评价的准确性和全面性。

参考文献：

[1] 新华网． 习近平：跟着总书记学习二十大报告丨科技是第一生产力[EB/OL]． [2022-11-23]． http://www.xinhuanet.com/politics/2022-11/203/c_1129152012.htm. (XINHUANET. XI J P: Follow the General Secretary to study the report of the 20th National Congress 丨 Science and technology is the first productive force [EB/OL]. [2022-11-23]. http://www.xinhuanet.com/politics/2022-11/203/c_1129152012.htm.)

[2] 中华人民共和国商务部. 中华人民共和国中央人民政府： 中共中央关于制定国民经济和社会发展第十四个五年规划和二〇三五年远景目标的建议[EB/OL]． [2020-11-03]． https://www.gov.cn/zhengce/2020-11/03/content_5556991.htm. (Ministry of Commerce of the People's Republic of China. Central People's Government

of the People's Republic of China: Proposal of the Central Committee of the Communist Party of China on the Formulation of the Fourteenth Five-Year Plan for National Economic and Social Development and the Visionary Goals for the 2035 Years [EB/OL]. [2020-11-03]. https://www.gov.cn/zhengce/2020-11/03/content_5556991.htm.

[3] SAUNILA M, UKKO J. A conceptual framework for the measurement of innovation capability and its effects[J]. Baltic Journal of Management, 2012, 7(4): 355-375.

[4] Cummings J L, TENG B S. Transferring R&D knowledge: the key factors affecting knowledge transfer success[J]. Journal of Engineering and technology management, 2003, 20(1-2): 39-68.

[5] MOED H F. Citation analysis in research evaluation[M]. Springer Science & Business Media, 2006.

[6] 李晶，杨雪，苏秋丹，等. 基于知识单元理论的科技成果创新性测度研究述评[J]. 现代情报，2023,43(8):161-177. (LI J, YANG X, SU Q D, et al. Review of the Research on the Measurement of Innovation of Scientific and Technological Achievements Based on the Theory of Knowledge Units [J]. Journal of Modern Information, 2023,43(8):161-177.)

[7] FLEMING L. Recombinant uncertainty in technological search[J]. Management science, 2001, 47(1): 117-132.

[8] LUO Z, LU W, HE J, et al．Combination of Research Questions and Methods: A New Measurement of Scientific Novelty[J]. Journal of Informetrics,2022,16(2):111282．

[9] POLAK M P, MORGAN D. Extracting accurate materials data from research papers with conversational language models and prompt engineering[J]. Nature Communications, 2024, 15(1): 1569.

[10] AMPLAYO R K， HONG S L， SONG M． Network－based Approach to Detect Novelty of Scholarly Literature[J]．Information Sciences, 2018, 422:542－557．

[11] VAN ECK N, WALTMAN L. Software survey: VOSviewer, a computer program for bibliometric map**[J]. Scientometrics, 2010, 84(2): 523-538.

[12] GONZALEZ-PEREIRA B, GUERRERO-BOTE V P, MOYA-ANEGON F. A new approach to the metric of journals' scientific prestige: The SJR indicator[J]. Journal of Informetrics, 2010, 4(3): 379-391.

[13] MOED H F. Measuring contextual citation impact of scientific journals[J]. Journal of Informetrics, 2010, 4(3): 265-277.

[14] PORTER A, RAFOLS I. Is science becoming more interdisciplinary? Measuring and map** six research fields over time[J]. Scientometrics, 2009, 81(3): 719-745.

[15] THOR A, MARX W, LEYDESDORFF L, et al. Introducing CitedReferencesExplorer (CRExplorer): A program for reference publication year spectroscopy with cited references standardization[J]. Journal of Informetrics, 2016, 10(2): 503-515.

[16] GAGOLEWSKI M. Bibliometric impact assessment with R and the CITAN package[J]. Journal of Informetrics, 2011, 5(4): 678-692.

[17] ALONSO S, CABRERIZO F, HERRERA-VIEDMA E, et al. hg-index: A new index to characterize the scientific output of researchers based on the h-and g-indices[J]. Scientometrics, 2010, 82(2): 391-400.

[18] SUD P, THELWALL M. Evaluating altmetrics[J]. Scientometrics, 2014, 98: 1131-1143.

[19] PRABOWO R, THELWALL M. Sentiment analysis: A combined approach[J]. Journal of Informetrics, 2009, 3(2): 143-157.

[20] HAMMARFELT B. Using altmetrics for assessing research impact in the humanities[J]. Scientometrics, 2014, 101(2): 1419-1430.

[21] VAN ZYL J M. A simulation study to investigate the accuracy of approximating averages of ratios using ratios of averages[J]. Journal of Informetrics, 2013, 7(4): 907-913.

[22] SIEDSCHLAG I, SMITH D, TURCU C, et al. What determines the location choice of R&D activities by multinational firms?[J]. Research Policy, 2013, 42(8): 1420-1430.

[23] PEPE A, Rodriguez M. Collaboration in sensor network research: an in-depth longitudinal analysis of assortative mixing patterns[J]. Scientometrics, 2010, 84(3): 687-701.

[24] BOUYSSOU D, Marchant T. An axiomatic approach to bibliometric rankings and indices[J]. Journal of Informetrics, 2014, 8(3): 449-477.
[25] SONG M, HEO G E, LEE D. Identifying the landscape of Alzheimer's disease research with network and content analysis[J]. Scientometrics, 2015, 102: 905-927.
[26] ZHENG J, ZHAO Z, ZHANG X, et al. Influences of counting methods on country rankings: a perspective from patent analysis[J]. Scientometrics, 2014, 98: 2087-2102.
[27] SCHREIBER M. Is the new citation-rank approach P100′ in bibliometrics really new?[J]. Journal of Informetrics, 2014, 8(4): 997-1004.
[28] FRANCESCHINI F, GALETTO M, MAISANO D, et al. An informetric model for the success-index[J]. Journal of Informetrics, 2013, 7(1): 109-116.
[29] WU C Y. Comparisons of technological innovation capabilities in the solar photovoltaic industries of Taiwan, China, and Korea[J]. Scientometrics, 2014, 98(1): 429-446.
[30] CHANG Y W. Exploring scientific articles contributed by industries in Taiwan[J]. Scientometrics, 2014, 99(2): 599-613.

(通信作者：陆凯余 Email:andre_lukaiyu@foxmail.com)

作者贡献说明/Author contributions：

陆凯余：选题，研究设计，数据收集与处理，论文撰写与修改。

杨雪：选题构思，论文修改。

A Study on the Identification Method of Paper Innovation Based on Mutation Index

Lu Kaiyu Yang Xue
School of Information Management, Sun Yat-sen University, Guangzhou 510006

Abstract: [Purpose/Significance] The aim is to construct a mutation index to measure the innovation of academic papers, providing a reference for improving the evaluation system of academic paper innovation. **[Method/Process]** A large language model (LLM), Mistral-openorca, is utilized to extract the research questions and methods from papers. The mutation index is constructed based on two dimensions: the similarity of research questions and methods to existing studies, and the extent to which they stimulate subsequent research. **[Result/Conclusion]** Papers with high innovation scores generally have a higher citation frequency; among the top 10 ranked papers by innovation, seven introduced original methods or tools, whereas papers in the bottom 10 ranked by innovation had more mature research topics and methods at the time of publication, with lower levels of innovation; the correlation coefficient between the innovation score of the papers and their citation frequency is 0.530, and empirical research results further confirm the effectiveness of the mutation index proposed in this paper in measuring the innovation of academic papers.

Keywords: Technological Innovation LLM Mutation Index Evaluation of Academic Paper Innovation

Authors: Lu Kaiyu, B.S. candidate, School of Information Management, Sun Yat-sen University, corresponding author E-mail: andre_lukaiyu@foxmail.com; Yang Xue, M.S. candidate, School of Information Management, Sun Yat-sen University.